APPARATUS AND METHOD FOR DETERMINING TRANSLATION WORD

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 USC 119(a) of Korean Patent Application No. 10-2015-0109154, filed on Jul. 31, 2015 in the Korean Intellectual Property Office, the entire disclosure of which is incorporated herein by reference for all purposes.

BACKGROUND

[0002] 1. Field

[0003] The following description relates to an automatic translation technology. The following description also relates to an apparatus for determining a translation word. The following description also relates to a method for determining a translation word.

[0004] 2. Description of Related Art

[0005] When trying to find a word with a similar meaning to a translation word, there exists a method for accomplishing this goal by using a synonym dictionary or a thesaurus to search for a translation word, such as by providing synonyms or words with similar meanings. Such a method operates by finding out a translation word with regard to the given words through providing a pair of synonyms or words with similar meanings, which are built in for each field. However, a great deal of effort and time by the professionals in the relevant field are required to build such a synonym dictionary or a thesaurus. In addition, even by using such a method, it is still difficult to find out the translation word with regard to new words. For example, which synonym is appropriate may depend upon the context of the translation.

[0006] Meanwhile, as machine learning technology develops, there is a method of finding out a translation word by learning a pair of parallel sentences through a parallel corpus in multiple languages. However, such an approach still results in requiring a lot of effort and time required for a task of building the corpus, where the pair of parallel words exists.

SUMMARY

[0007] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0008] In one general aspect, an apparatus for determining a translation word includes a word vector generator configured to generate a word vector corresponding to an input word of a first language with reference to a first word vector space that is related to the first language, a word vector determiner configured to determine a word vector of a second language, wherein the determined word vector of the second language corresponds to the generated word vector, using a matching model, and a translation word selector configured to select a translation word of the second language, wherein the selected translation word corresponds to the input word of the first language, based on the determined word vector of the second language.

[0009] The apparatus may further include a word inputter configured to receive the input word of the first language.

[0010] The translation word selector may be configured to select a word vector, wherein the selected word vector is the most similar to the determined word vector of the second language, from among word vectors on a second word vector space that is related to the second language, and select, as the translation word, a word of the second language, wherein the word of the second language corresponds to the selected word vector.

[0011] The translation word selector may be configured to select the word vector, wherein the selected word vector is the most similar to the determined word vector of the second language, from among the word vectors on the second word vector space by using at least one of a distance measurement function, a similarity measurement function, or a correlation coefficient.

[0012] The distance measurement function may be one of Euclidean distance, Mahalanobis distance, or Hamming distance, the similarity measurement function may be cosine similarity, and the correlation coefficient may be one of Pearson correlation coefficient, Spearman correlation coefficient, partial correlation coefficient, or Cronbach's alpha.

[0013] The first word vector space may be built in advance through using machine learning using a first language corpus, and the second word vector space may be built in advance using machine learning using a second language corpus.

[0014] The first word vector space may be built by generating word vectors for each word of the first language and mapping the generated word vectors for each word of the first language to the vector space of the first language, and the second word vector space may be built by generating word vectors for each word of the second language and mapping the generated word vectors for each word of the second language to the vector space of the second language.

[0015] A machine learning algorithm used during machine learning may be one of a neural network, a decision tree, a genetic algorithm (GA), a genetic programming (GP), a Gaussian process regression, a linear discriminant analysis (LDA), a K-near neighbor (K-NN), a perceptron algorithm, a radial basis function network, a support vector machine (SVM), and deep-learning.

[0016] The matching model may be a model for matching the first word vector space to a second word vector space that is related to the second language.

[0017] The matching model may be built in advance through machine learning by using language resources that define a relation between the first and second languages.

[0018] The language resources may include at least one of a synonym dictionary or a thesaurus.

[0019] The determined translation word may not be predefined as a translation by the language resources.

[0020] In another general aspect, a method of determining a translation word includes generating a word vector corresponding to an input word of a first language with reference to a first word vector space that is related to the first language, determining a word vector of a second language, wherein the determined word vector of the second language corresponds to the generated word vector, by using a matching model, and selecting a translation word of the second language, wherein the selected transition word corresponds to the input word of the first language, based on the determined word vector of the second language.

[0021] The method may further include receiving the input word of the first language.